Causality 101

How to draw causal statements from data and get away with it

Lizard Lab - 09.02.2024

UNIVERSITEIT GENT

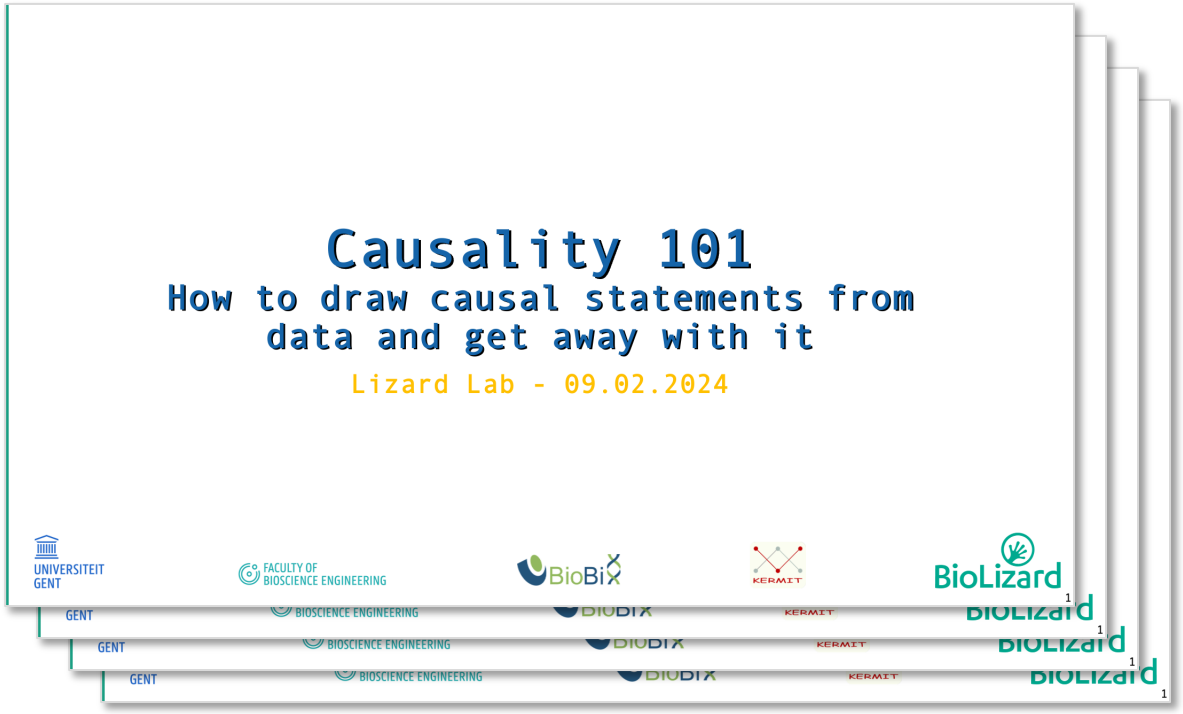
FACULTY OF BIOSCIENCE ENGINEERING

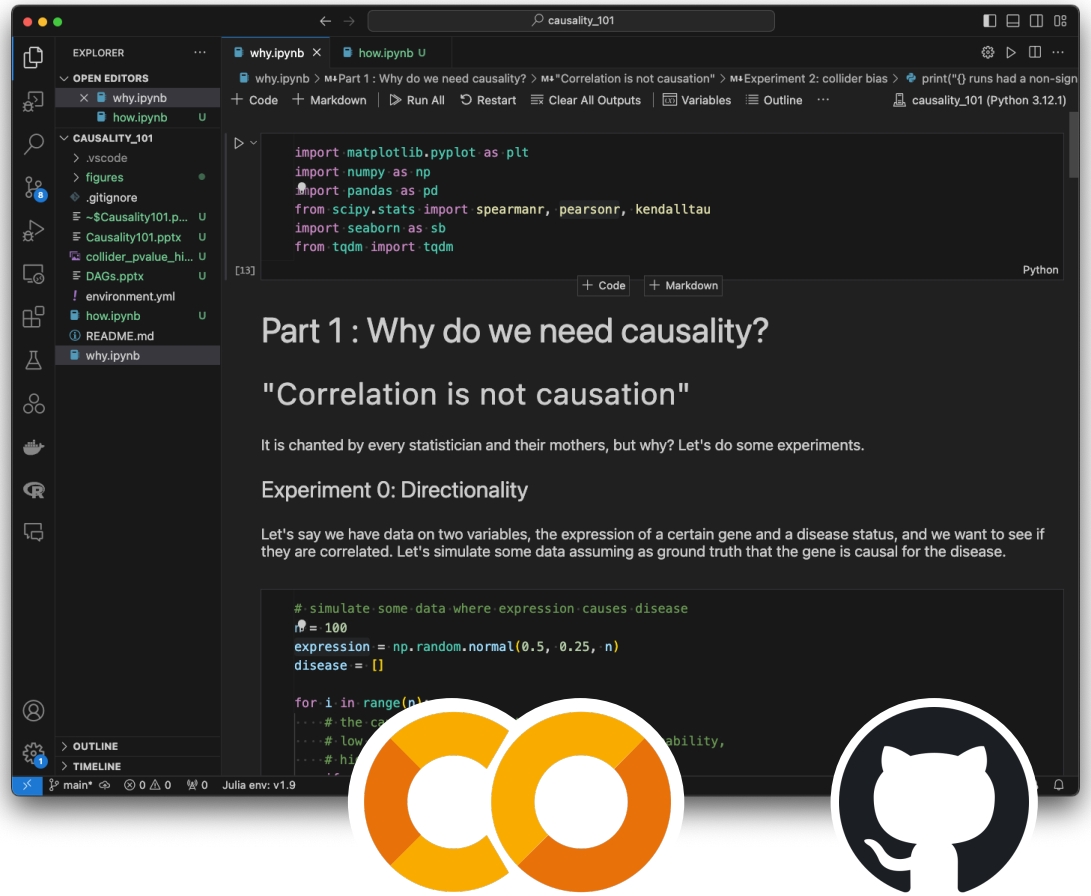
BioBix

KERMIT

BioLizard

Structure of the session





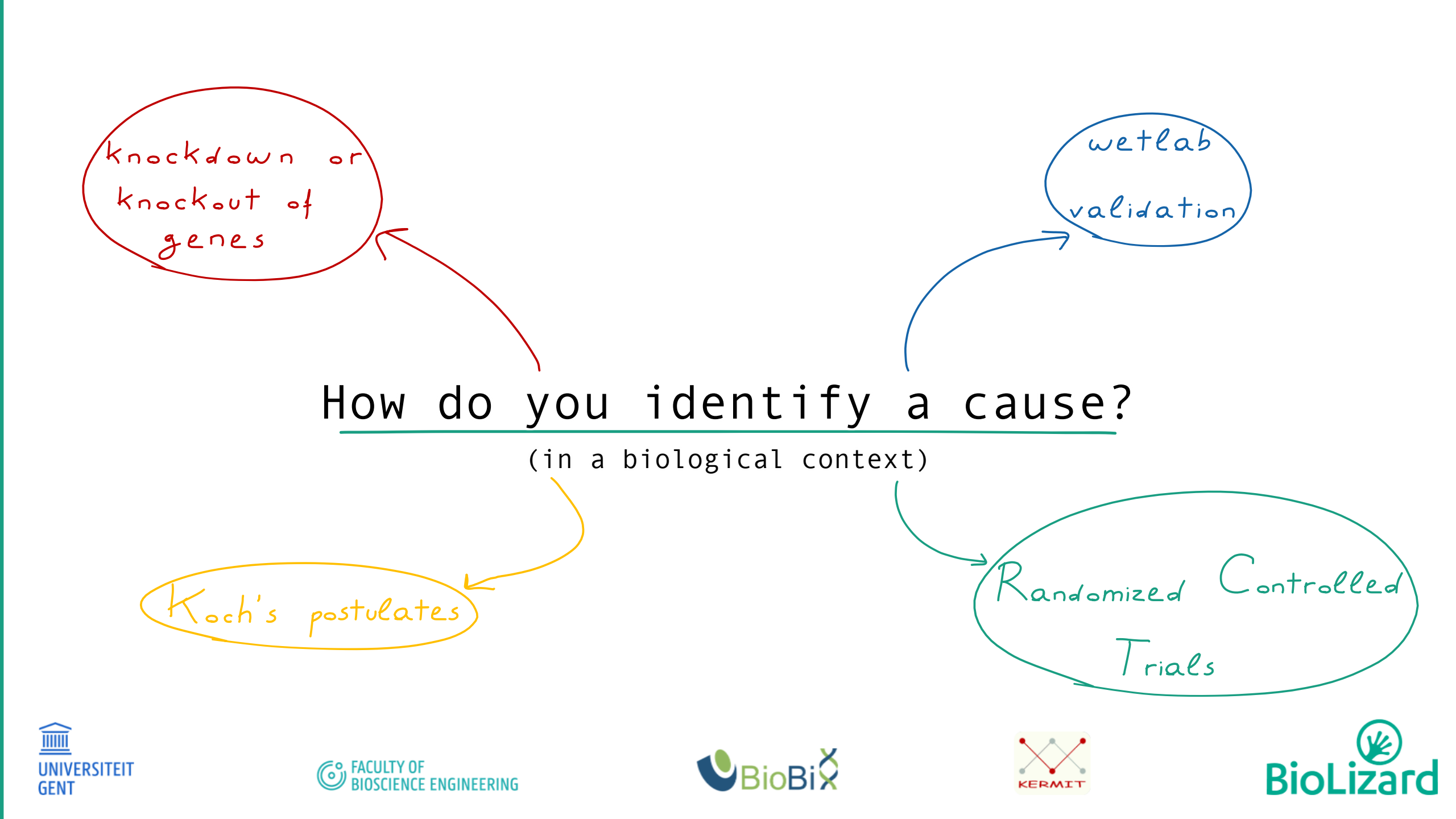
UNIVERSITEIT GENT

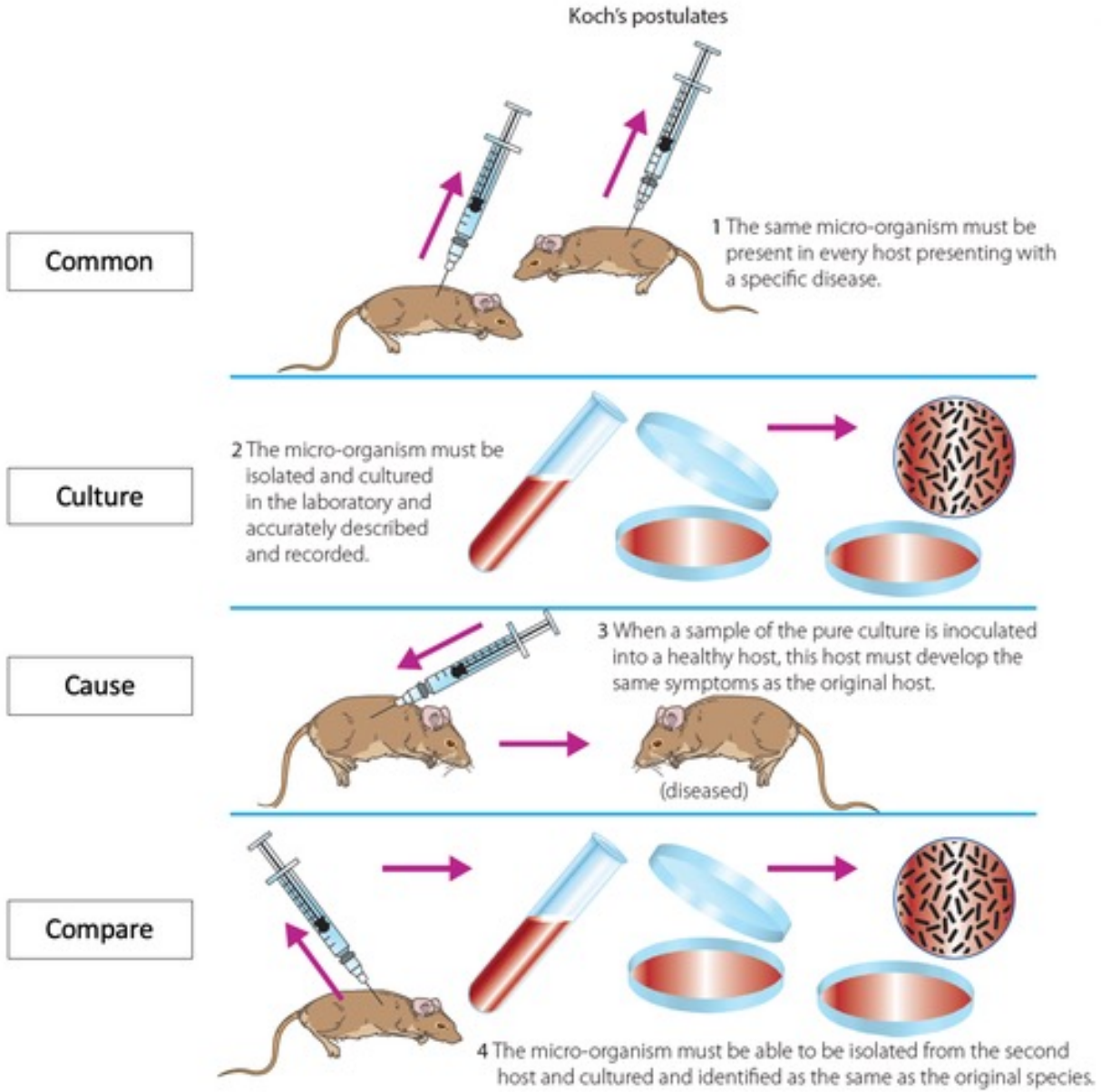
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

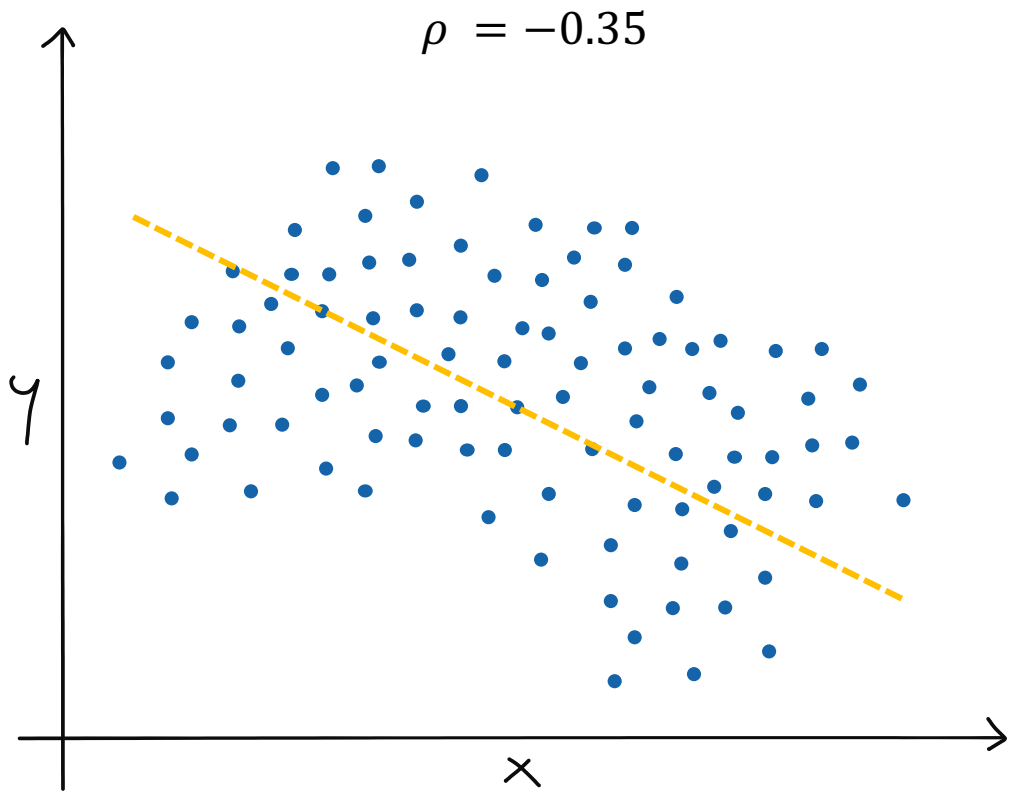
KERMIT

BioLizard

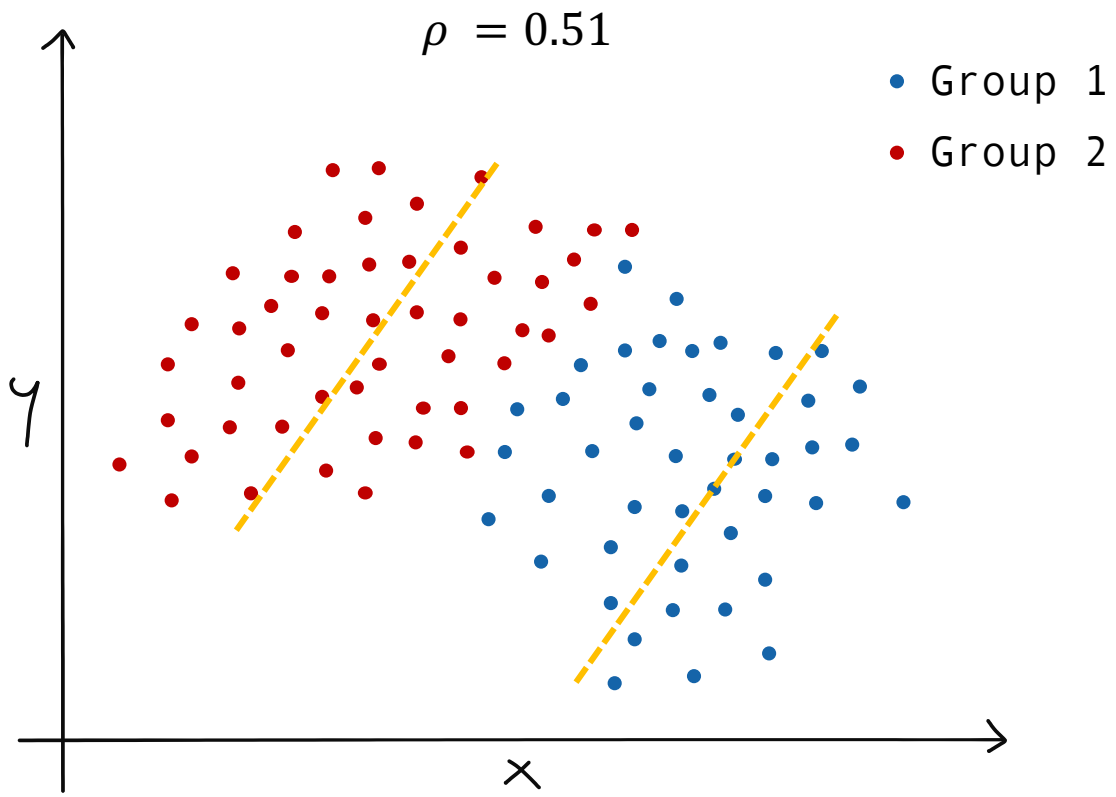

knockdown or knockout of genes
wetlab validation
How do you identify a cause?
(in a biological context)
Koch's postulates
Randomized Controlled Trials


Koch's postulates
Common
1 The same micro-organism must be present in every host presenting with a specific disease.
Culture
2 The micro-organism must be isolated and cultured in the laboratory and accurately described and recorded.
Cause
3 When a sample of the pure culture is inoculated into a healthy host, this host must develop the same symptoms as the original host.
(diseased)
Compare
4 The micro-organism must be able to be isolated from the second host and cultured and identified as the same as the original species.

Why correlation ≠ causation

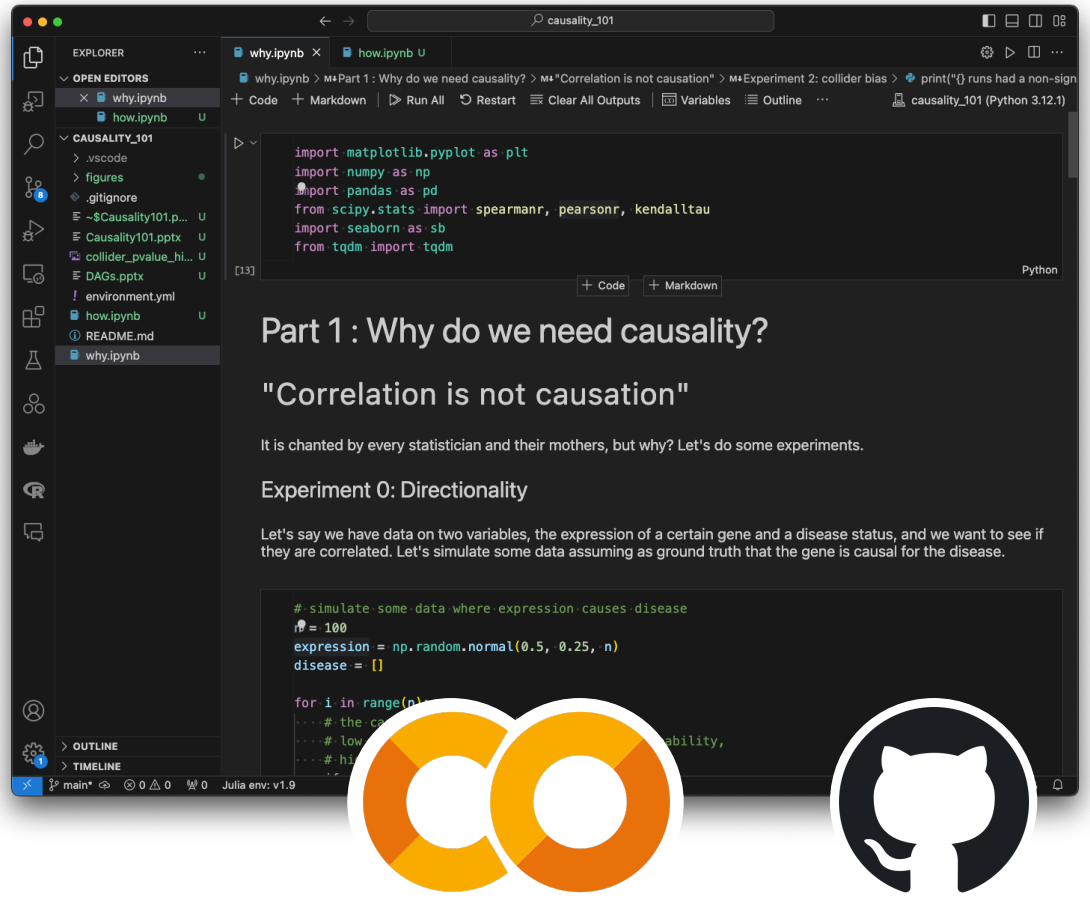


Why correlation ≠ causation



Why correlation ≠ causation

why.ipynb



UNIVERSITEIT GENT

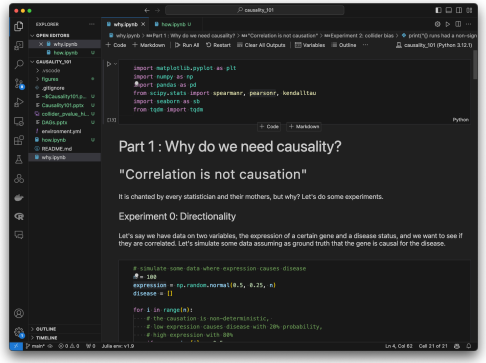
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

KERMIT

BioLizard

Why correlation ≠ causation



UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

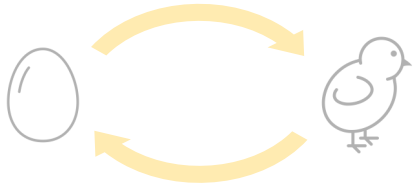
BioBix

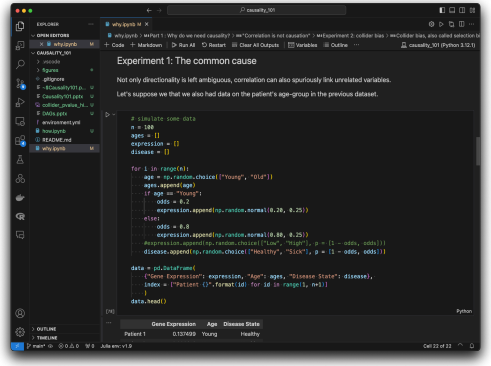
KERMIT

BioLizard

Why correlation ≠ causation

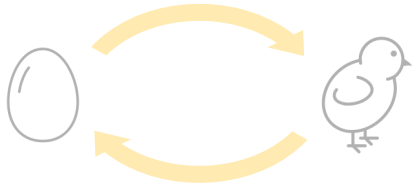
- directionality

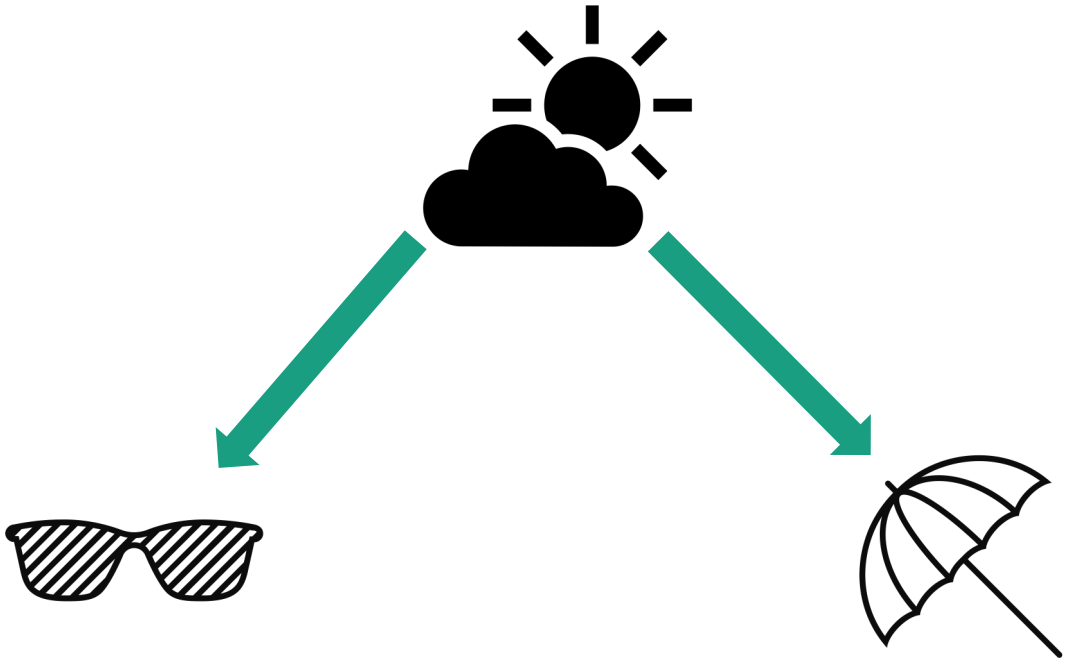


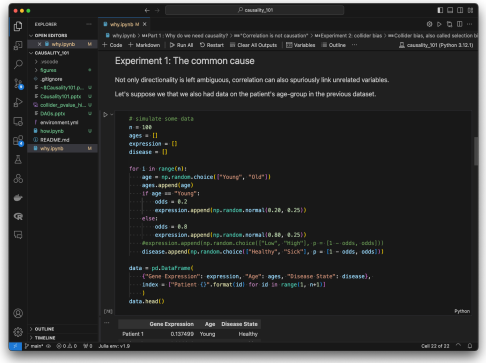


Why correlation ≠ causation

- directionality

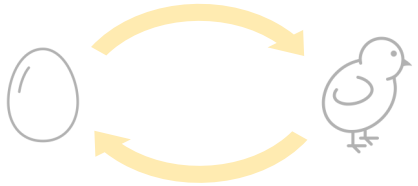




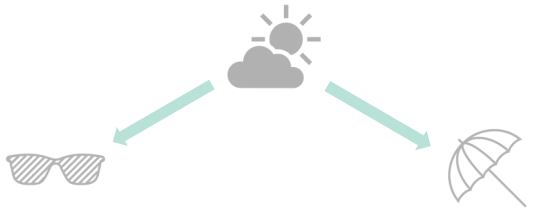


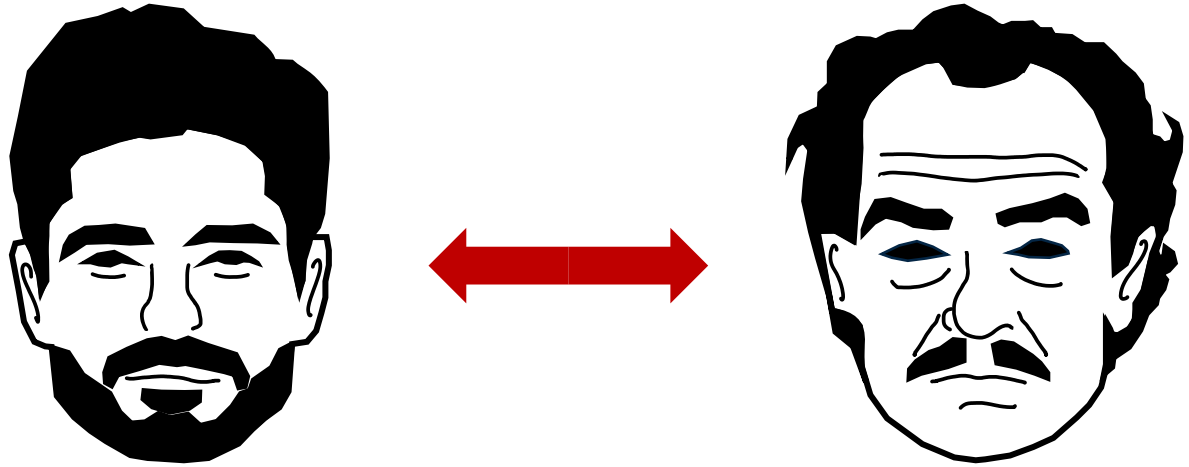
Why correlation ≠ causation

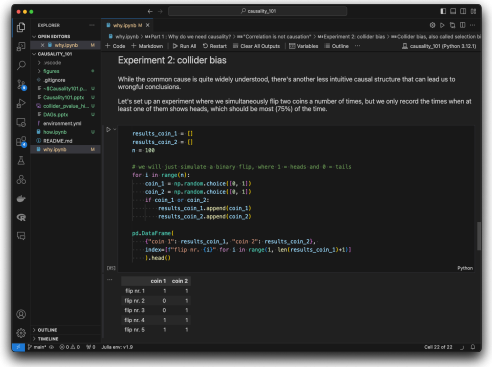
- directionality



- common causes







UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

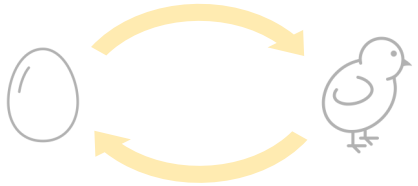
BioBix

KERMIT

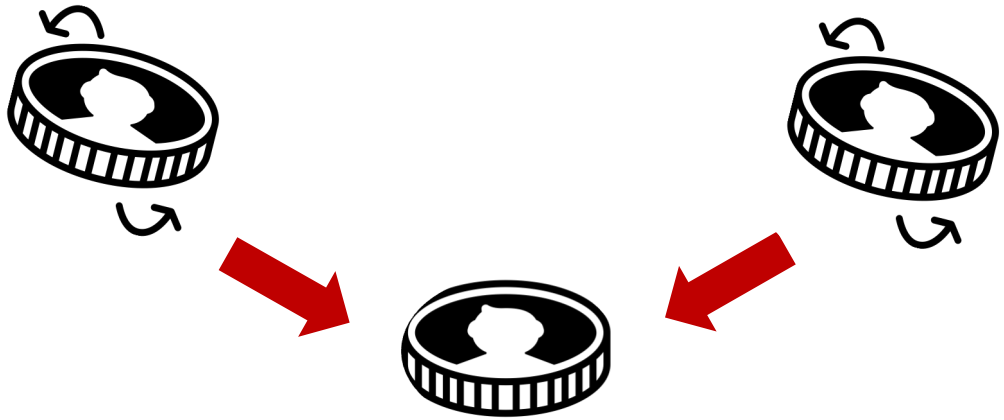
BioLizard

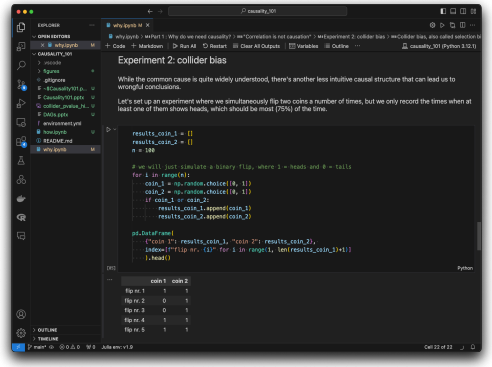
Why correlation ≠ causation

- directionality



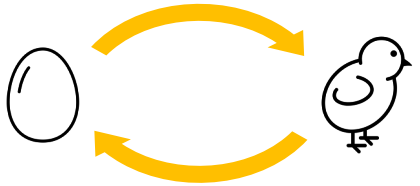
- common causes



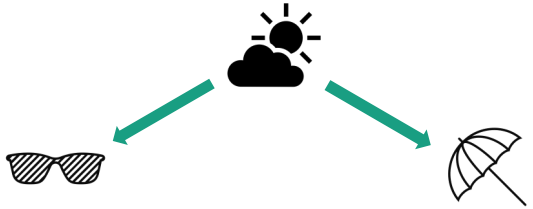


Why correlation ≠ causation

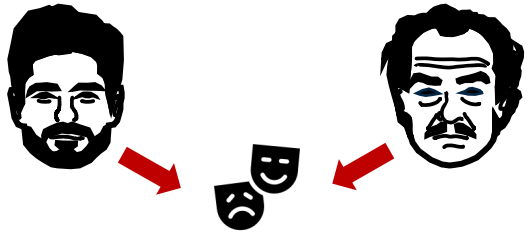
- directionality

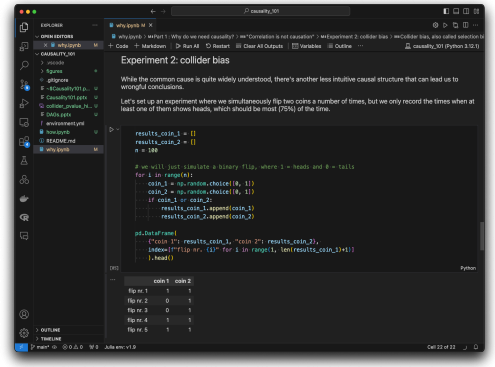


- common causes

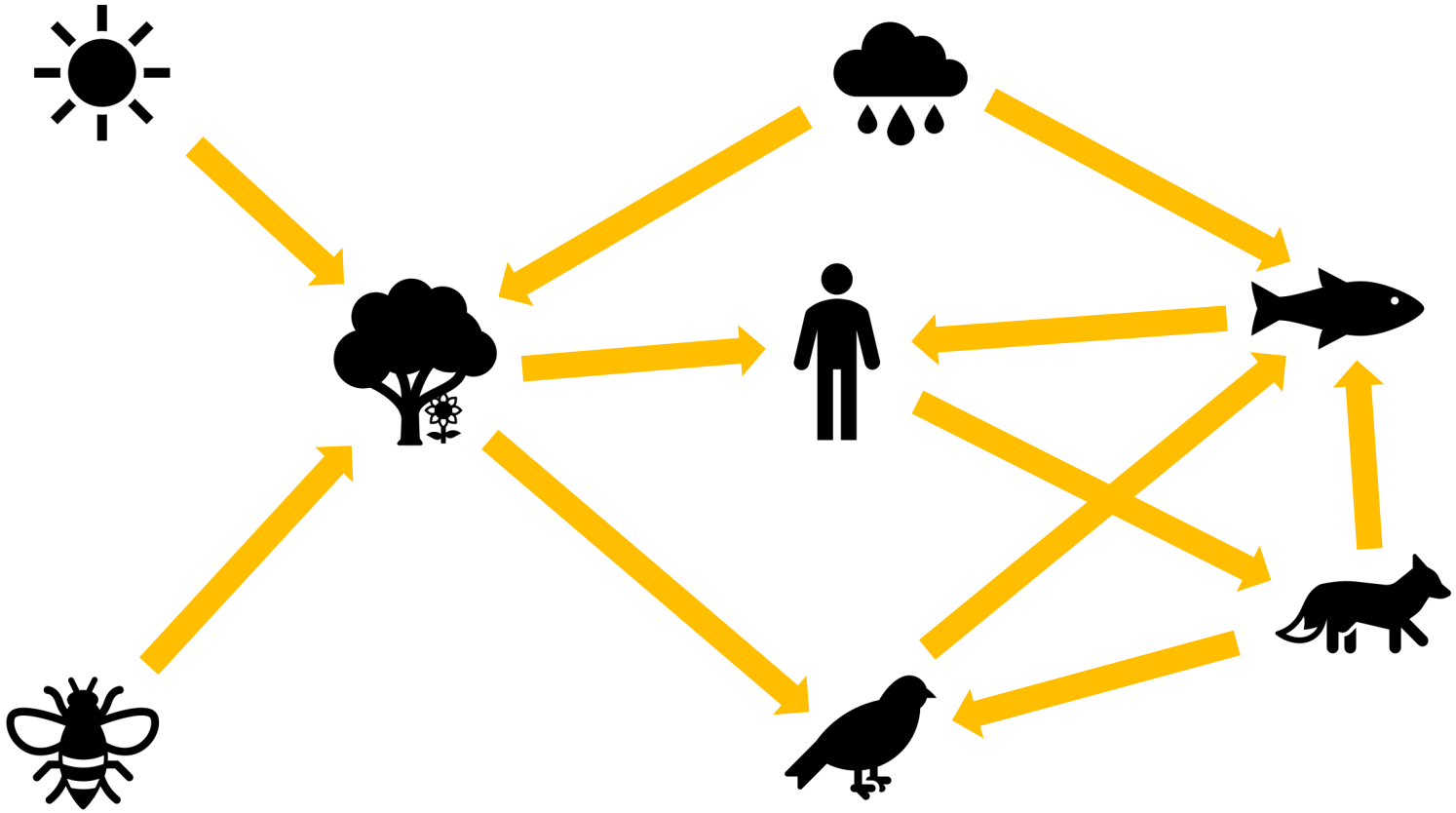


- collider/selection bias





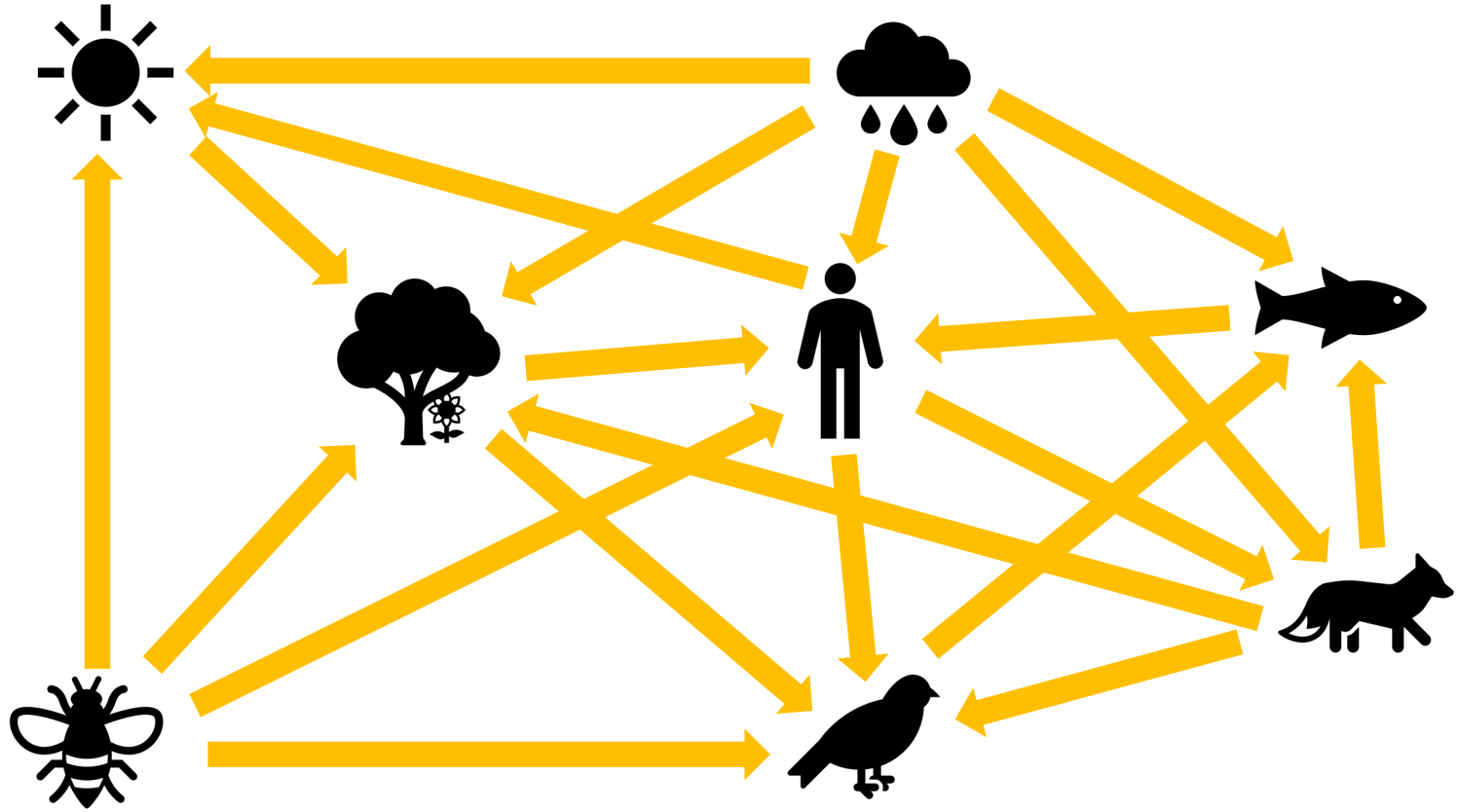
How do we deal with causality?



Causal Directed Acyclic Graphs (DAGs)

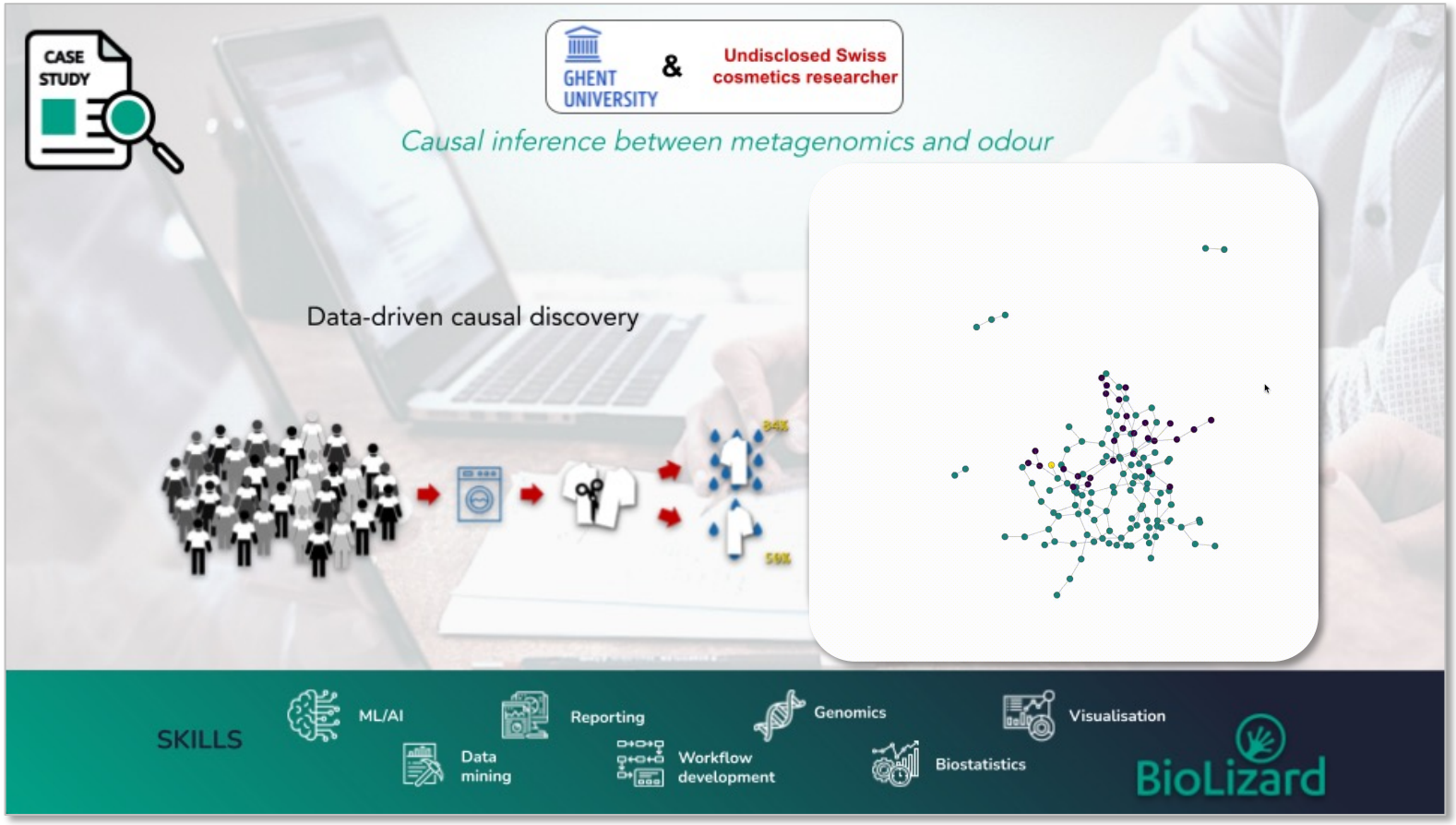
Intermezzo – causal discovery

obs.	☀	🌧	🐟	🧍	🌳	🐦
1	17°	1.8	19	100	42	12
2	22°	2.2	20	105	44	14
3	15°	0.9	21	106	39	8
4	18°	1.2	18	104	40	10
5	24°	0	17	100	41	11
6	19°	1.5	18	99	42	13
7	20°	1.6	18	103	58	7



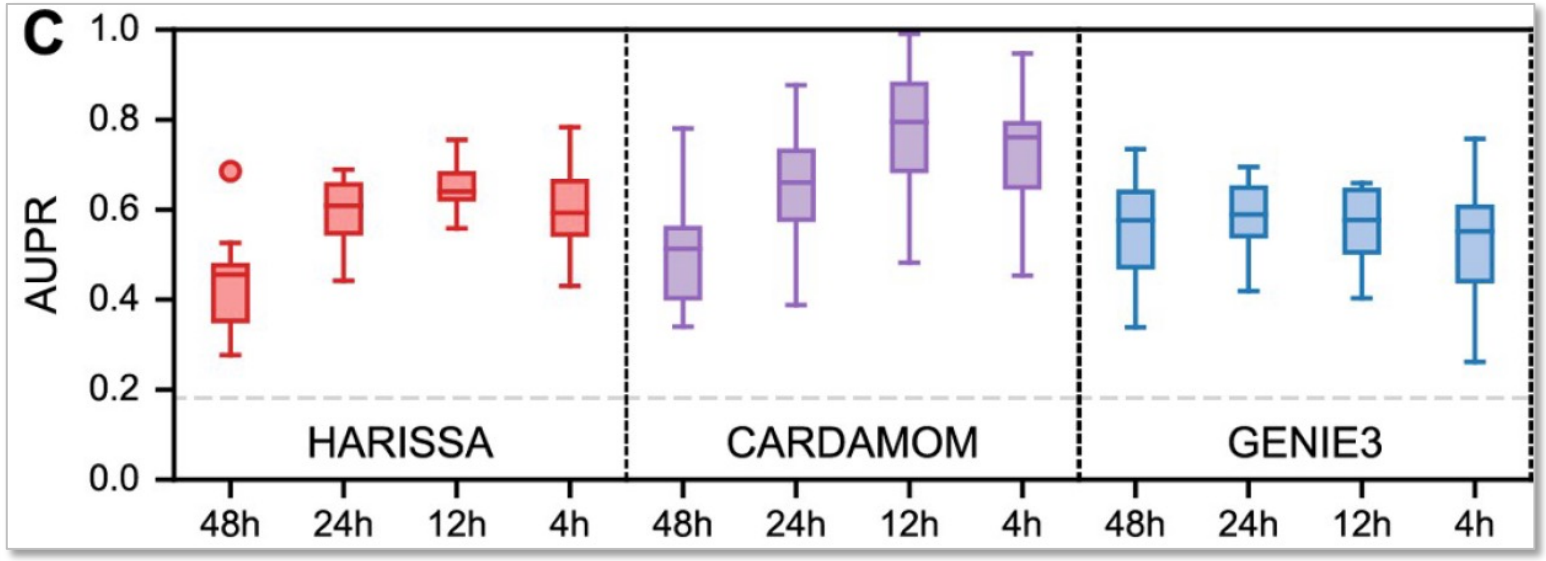
Intermezzo - causal discovery

Intermezzo – causal discovery



Intermezzo – causal discovery

Interesting field, but state-of-the-art is still not very accurate



[Submitted on 29 Aug 2023]

The CausalBench challenge: A machine learning contest for gene network inference from single-cell perturbation data

Mathieu Chevalley, Jacob Sackett-Sanders, Yusuf Roohani, Pascal Notin, Artemy Bakulin, Dariusz Brzezinski, Kaiwen Deng, Yuanfang Guan, Justin Hong, Michael Ibrahim, Wojciech Kotlowski, Marcin Kowiel, Panagiotis Misiakos, Achille Nazaret, Markus Püschel, Chris Wendler, Arash Mehrjou, Patrick Schwab

Nonlinear causal discovery with additive noise models

Part of Advances in Neural Information Processing Systems 21 (NIPS 2008)

Bibtex | Metadata | Paper

Authors

Patrik Hoyer, Dominik Janzing, Joris M.

UNIVERSITEIT GENT

Learning Functional Causal Models with Generative Neural Networks

Olivier Goudet, Diviyan Kalainathan, Philippe Caillou, Isabelle Guyon, David Lopez-Paz & Michè Sebag

Chapter | First Online: 30 November 2018

3776 Accesses | 12 Citations | 1 Altmetric

Conditional Distribution Variability Measures for Causality Detection

Josè A. R. Fonollosa

Chapter | First Online: 23 October 2019

658 Accesses | 1 Citations

Article | Open access | Published: 26 October 2023

Deep learning of causal structures in high dimensions under data limitations

Kai Lagemann, Christian Lagemann, Bernd Taschler & Sach Mukherjee

Nature Machine Intelligence 5, 1306–1316 (2023) | Cite this article

Analysis of cause-effect inference by comparing regression errors

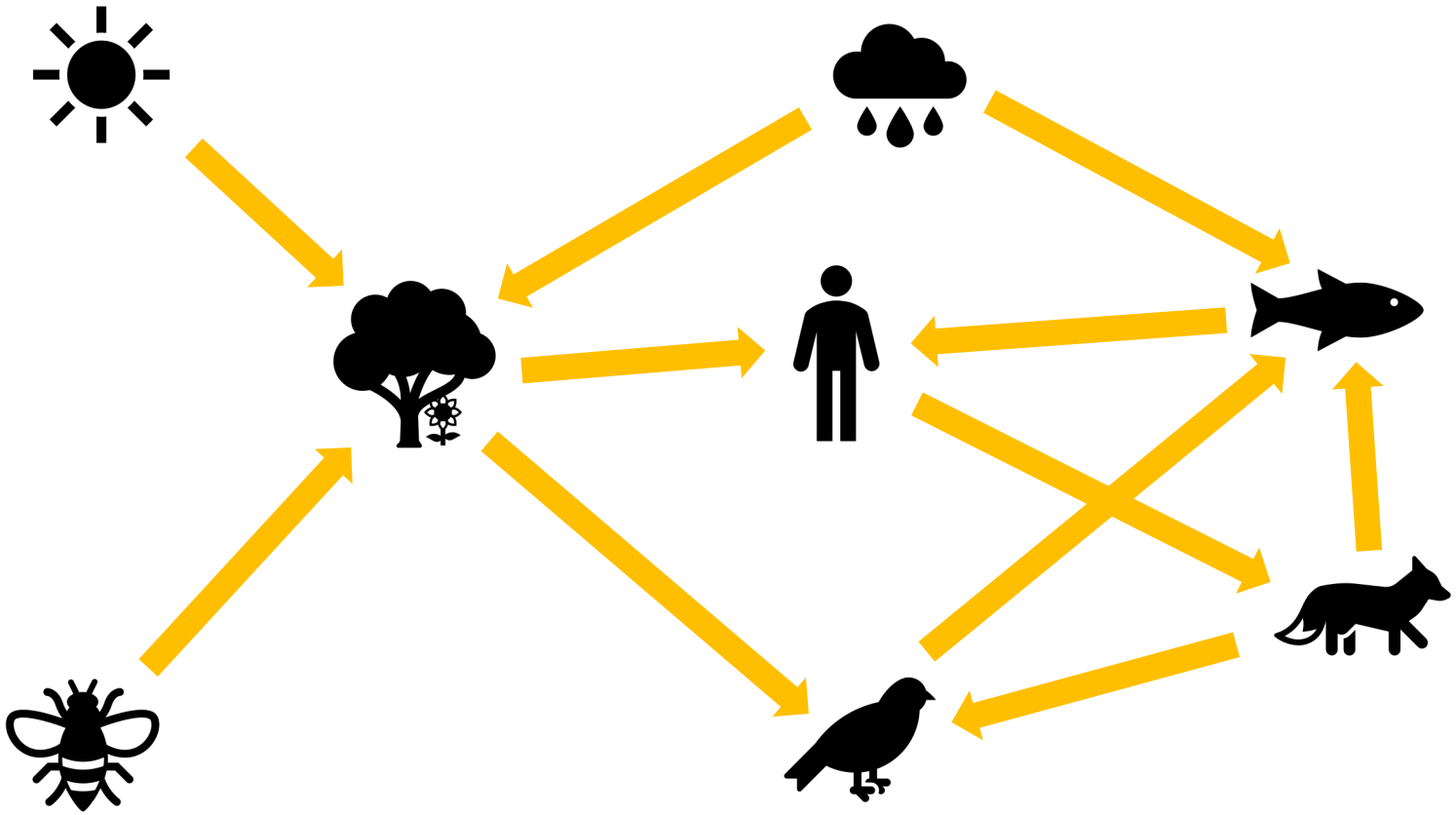
Research article | Artificial Intelligence | Data Mining and Machine Learning

Patrick Blöbaum[1], Dominik Janzing[2], Takashi Washio[1], Shohei Shimizu[3], Bernhard Schölkopf[2]

January 21, 2019

BioLizard

Intermezzo - causal inference



 →  : $p = 0.007$

If  x2, then  x1.2

Had  x2, but  =,

then  x0.8

UNIVERSITEIT GENT

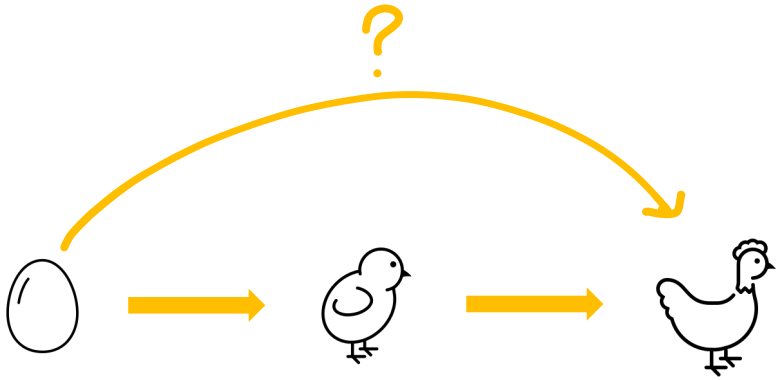
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

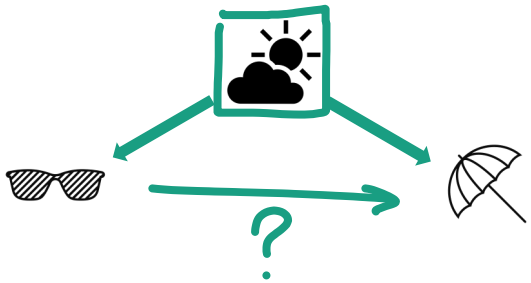
KERMIT

BioLizard

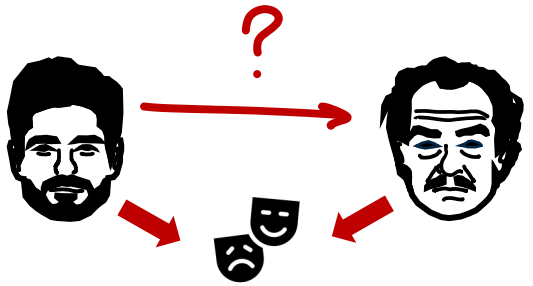
How do we deal with causality?



Don't adjust!

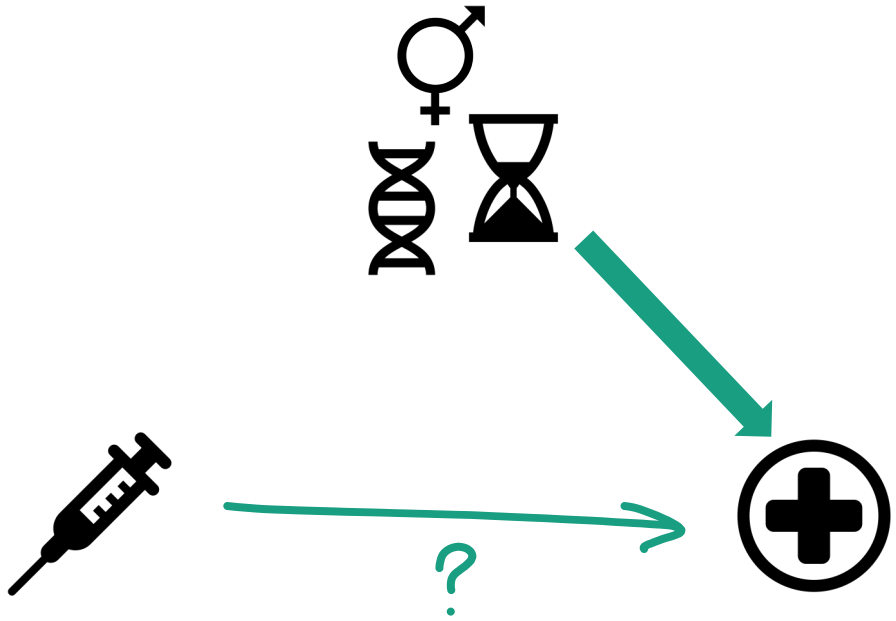


Adjust!



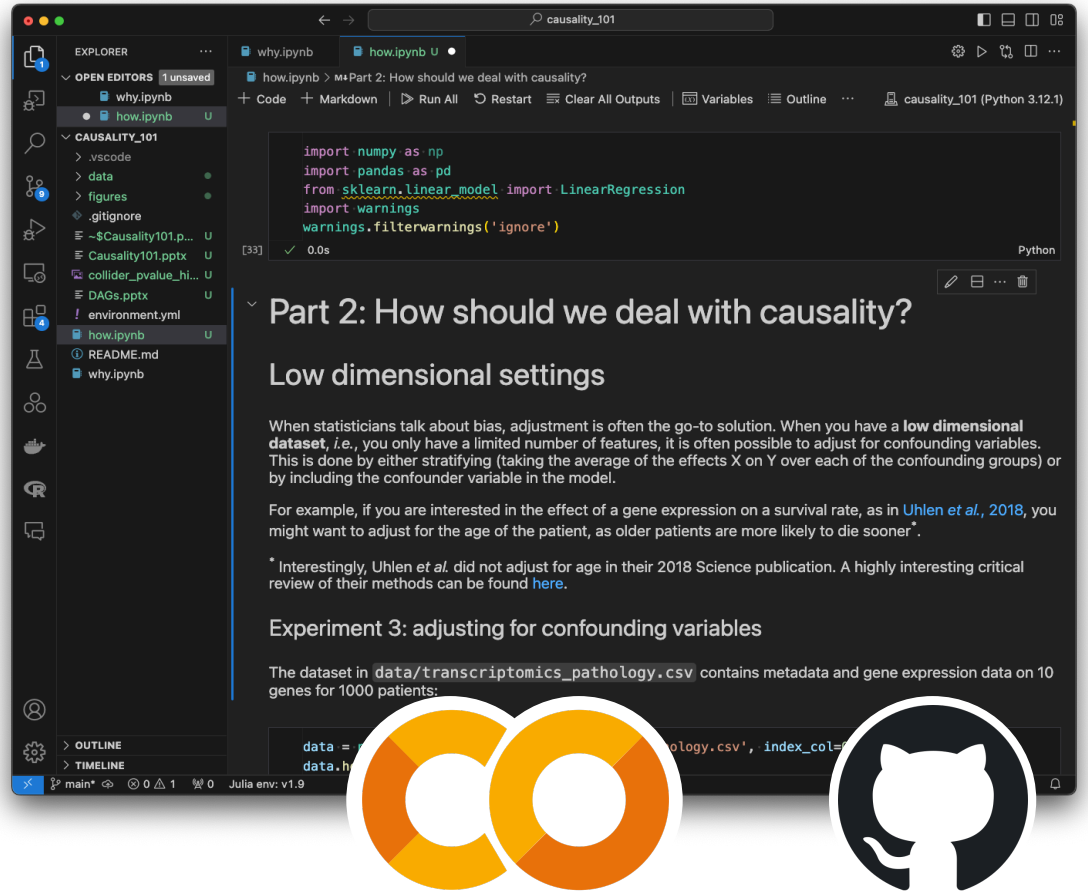
Don't adjust!

Why do RCTs work?



Causal inference

how.ipynb



UNIVERSITEIT GENT

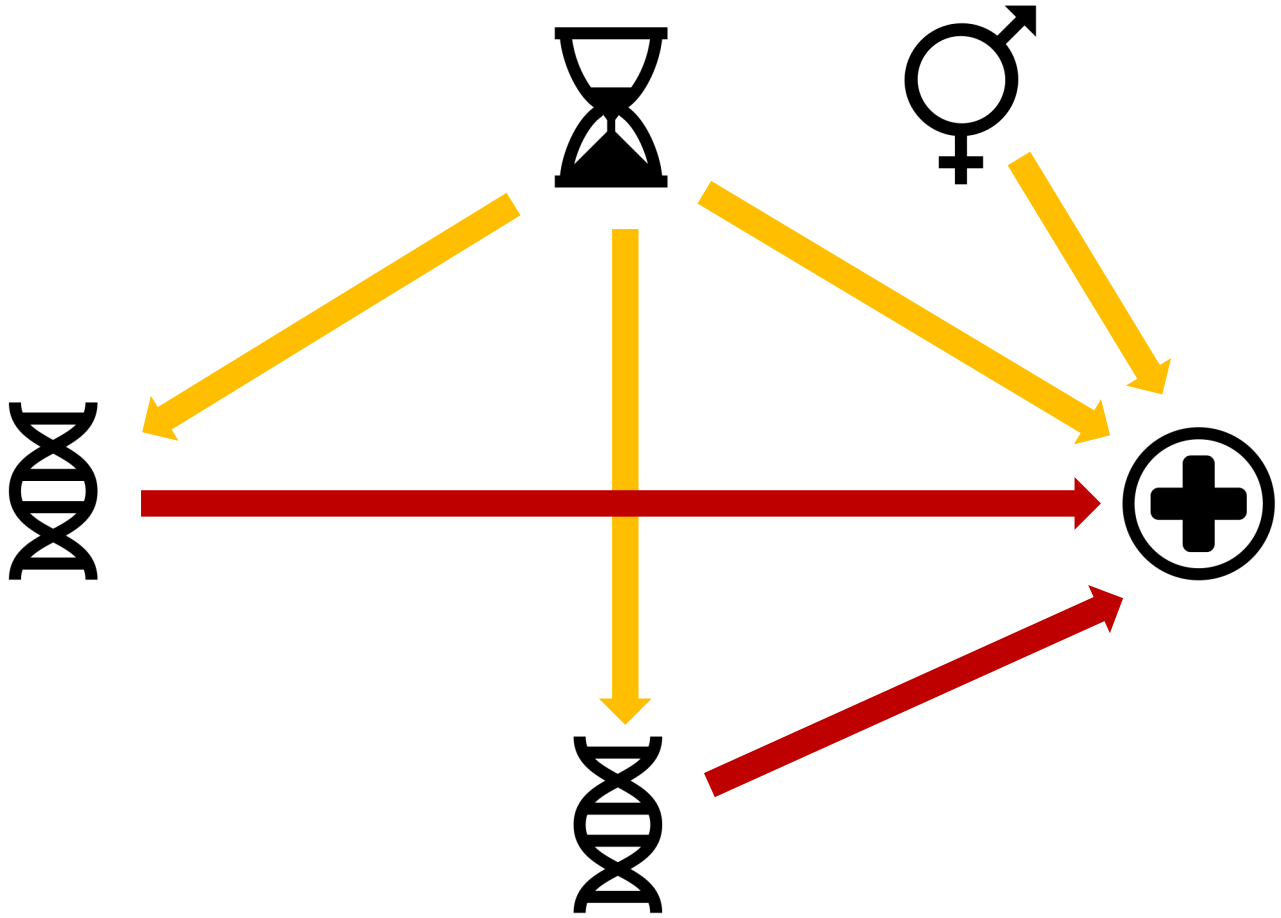
FACULTY OF BIOSCIENCE ENGINEERING

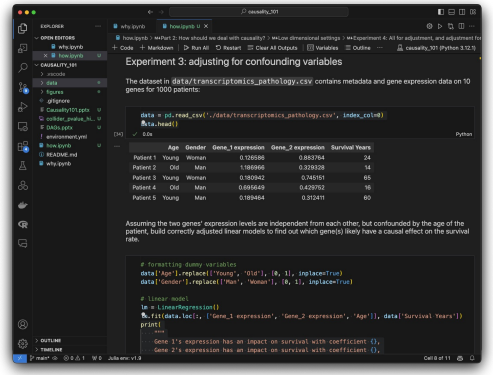
BioBix

KERMIT

BioLizard

Causal inference





UNIVERSITEIT GENT

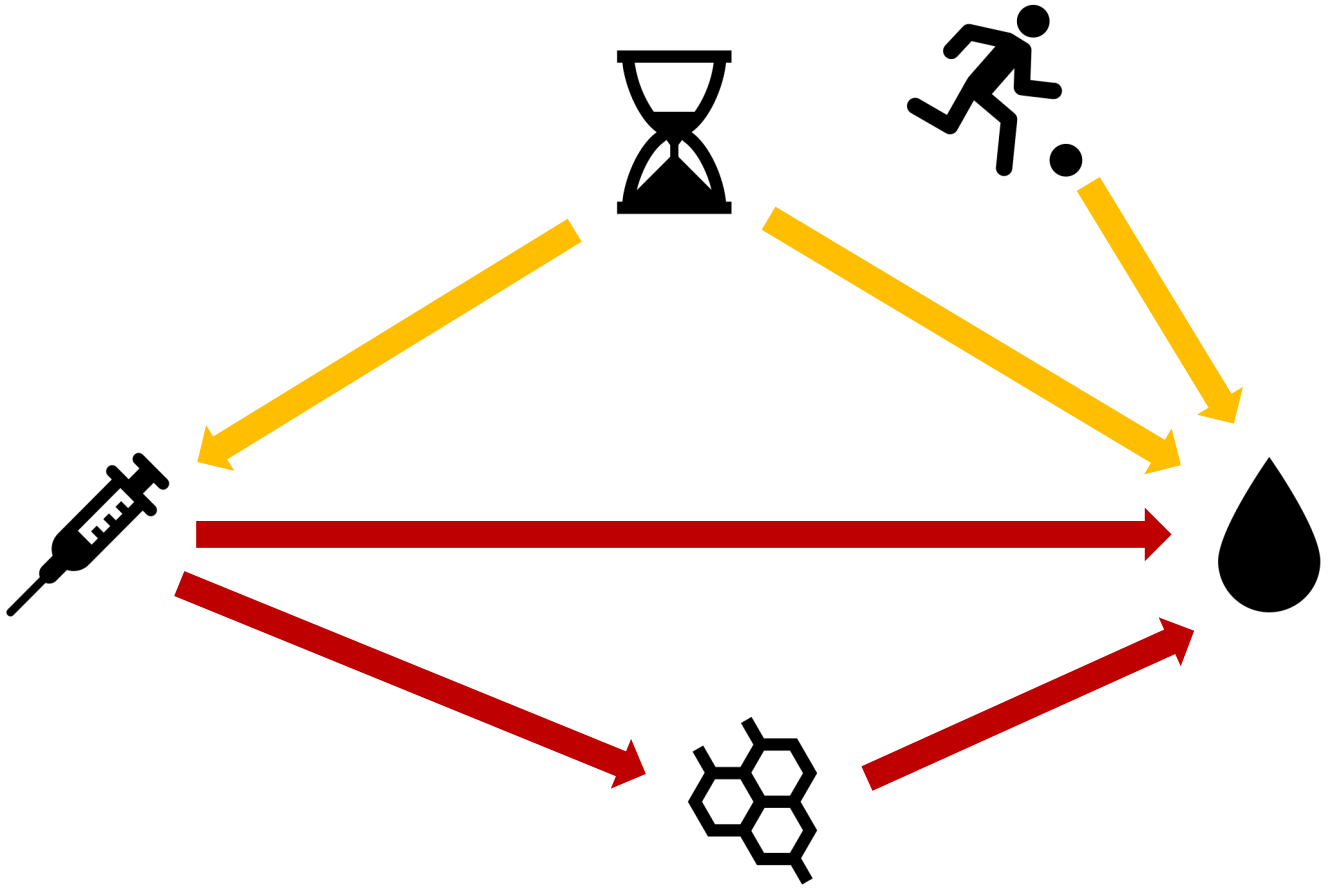
FACULTY OF BIOSCIENCE ENGINEERING

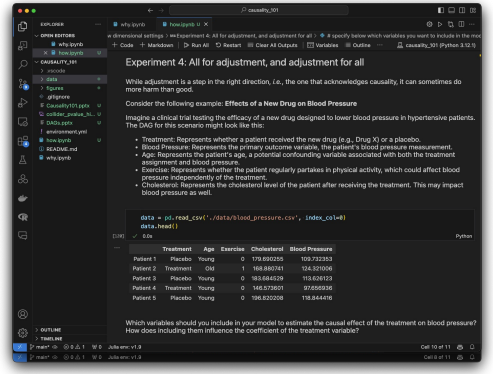
BioBix

KERMIT

BioLizard

Causal inference





UNIVERSITEIT GENT

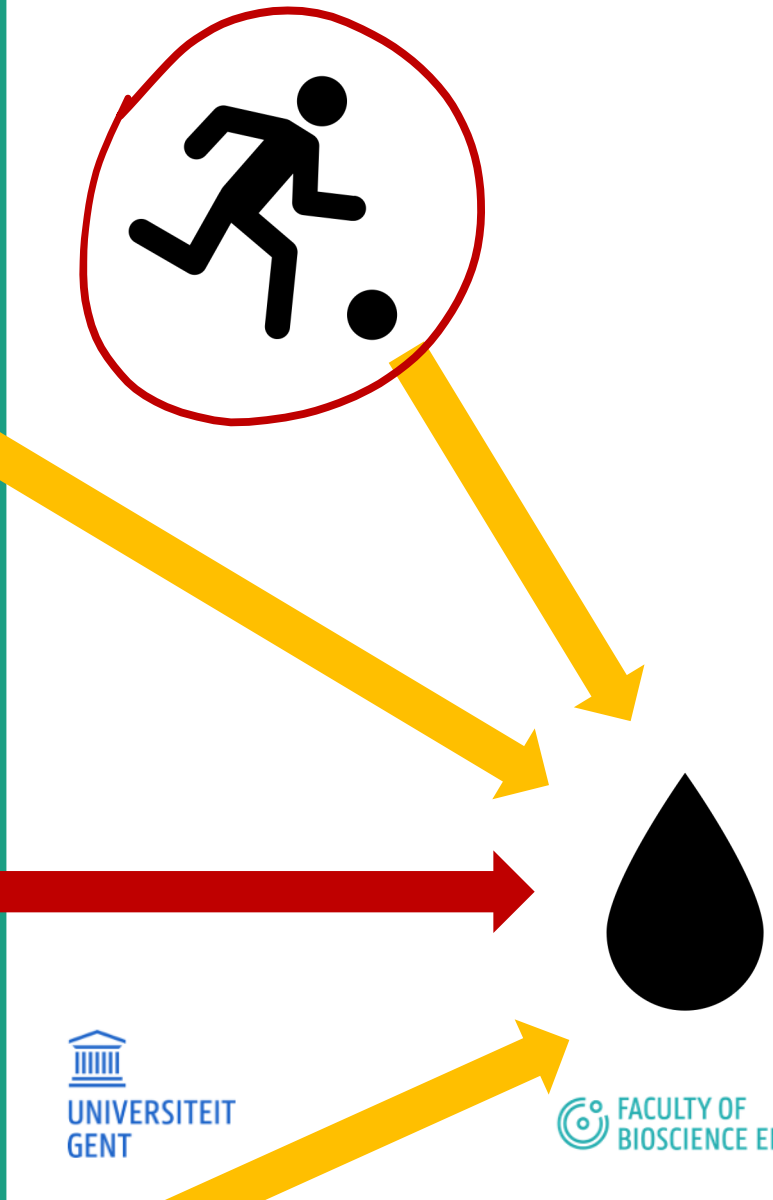
FACULTY OF BIOSCIENCE ENGINEERING

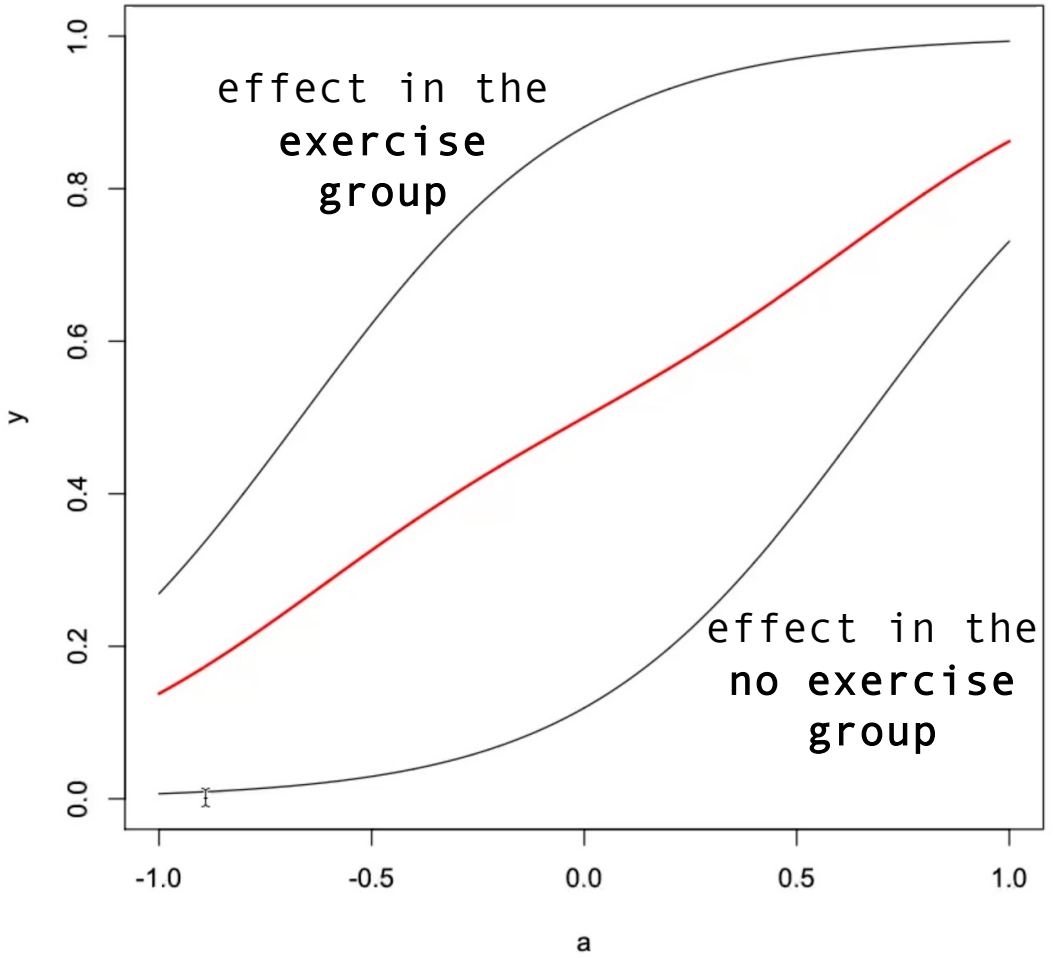
BioBix

KERMIT

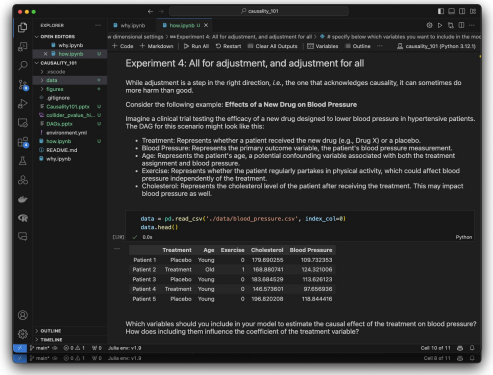
BioLizard

Should you adjust for exercise?





You can but don't have to



UNIVERSITEIT GENT

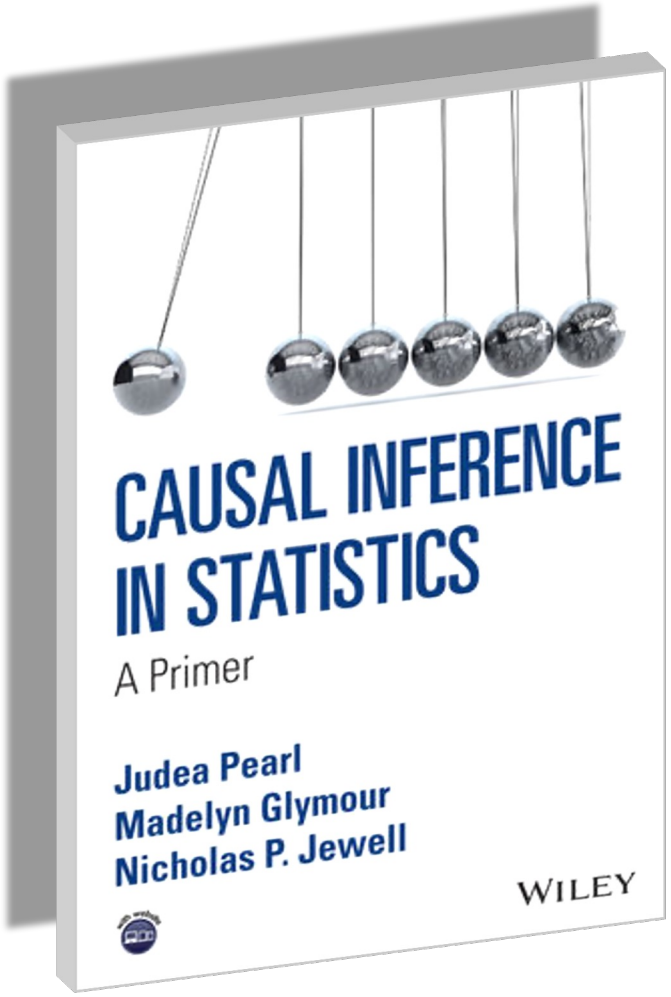
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

KERMIT

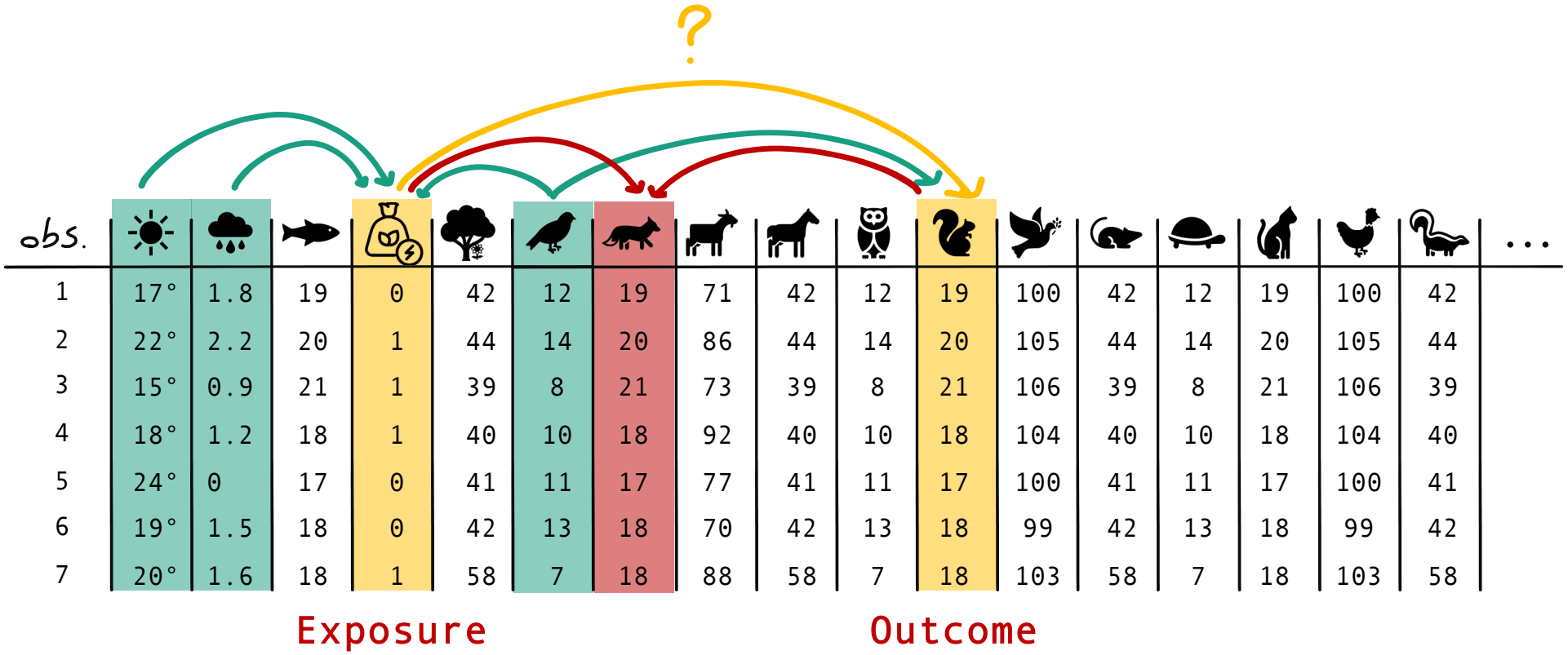
BioLizard

More about causal inference*



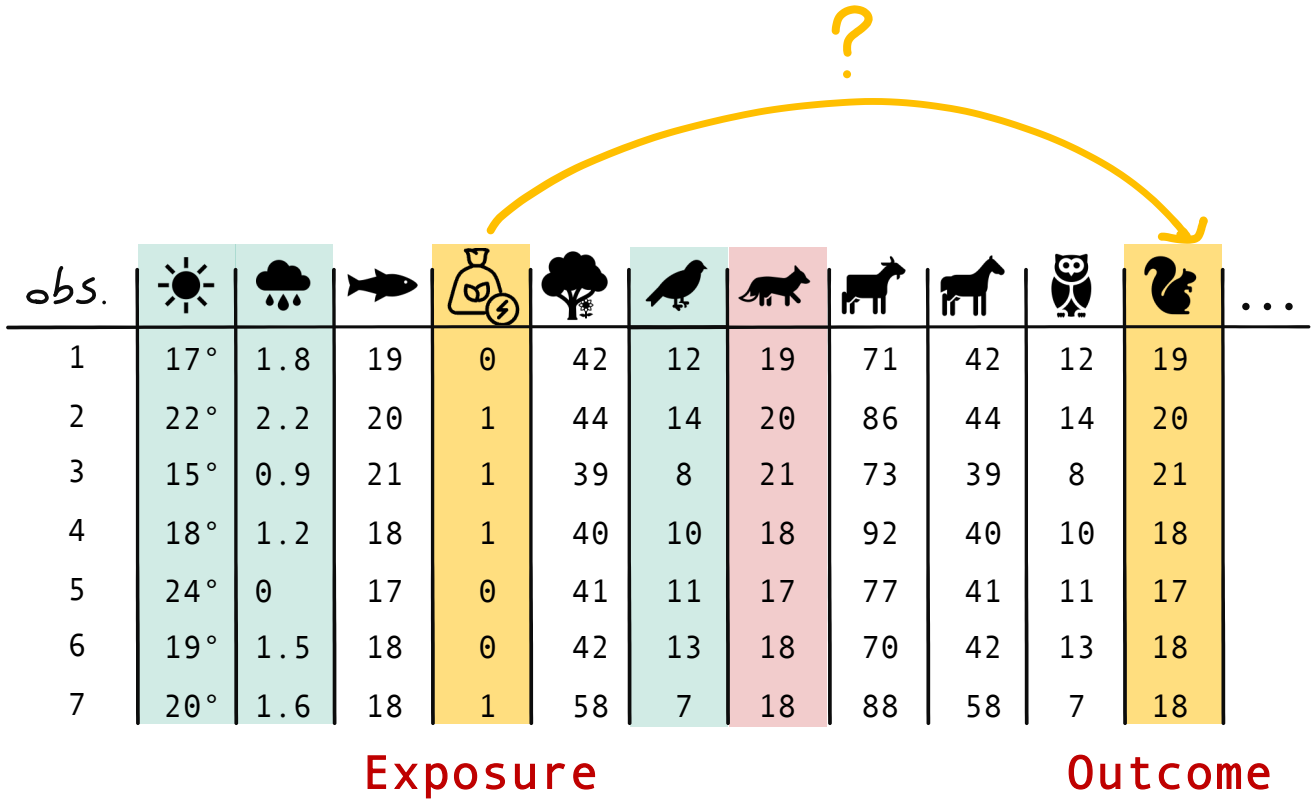
*in low dimensional data settings

What about high dimensional data?



obs.																		...
1	17°	1.8	19	0	42	12	19	71	42	12	19	100	42	12	19	100	42	
2	22°	2.2	20	1	44	14	20	86	44	14	20	105	44	14	20	105	44	
3	15°	0.9	21	1	39	8	21	73	39	8	21	106	39	8	21	106	39	
4	18°	1.2	18	1	40	10	18	92	40	10	18	104	40	10	18	104	40	
5	24°	0	17	0	41	11	17	77	41	11	17	100	41	11	17	100	41	
6	19°	1.5	18	0	42	13	18	70	42	13	18	99	42	13	18	99	42	
7	20°	1.6	18	1	58	7	18	88	58	7	18	103	58	7	18	103	58	

Causal machine learning



Propensity model:

$$\hat{E}(\text{bag}=1|\text{sun},\ \text{cloud},\ \text{bird},\ \ldots) \rightarrow g(L)$$

Outcome model:

$$\hat{E}(\text{squirrel}\ |\ \text{bag}=0,\ \text{sun},\ \text{cloud},\ \text{bird},\ \ldots) \rightarrow Q^{(0)}(L)$$

$$E[\text{squirrel}^1 - \text{squirrel}^0] = \frac{\sum_{i=1}^{n} \text{bag}_i(\text{squirrel}_i - Q^{(0)}(L_i))}{\sum_{i=1}^{n} \text{bag}_i} - \sum_{i=1}^{n}(1-\text{bag}_i)\frac{g(L_i)}{1-g(L_i)}(\text{squirrel}_i - Q^{(0)}(L_i))$$

UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

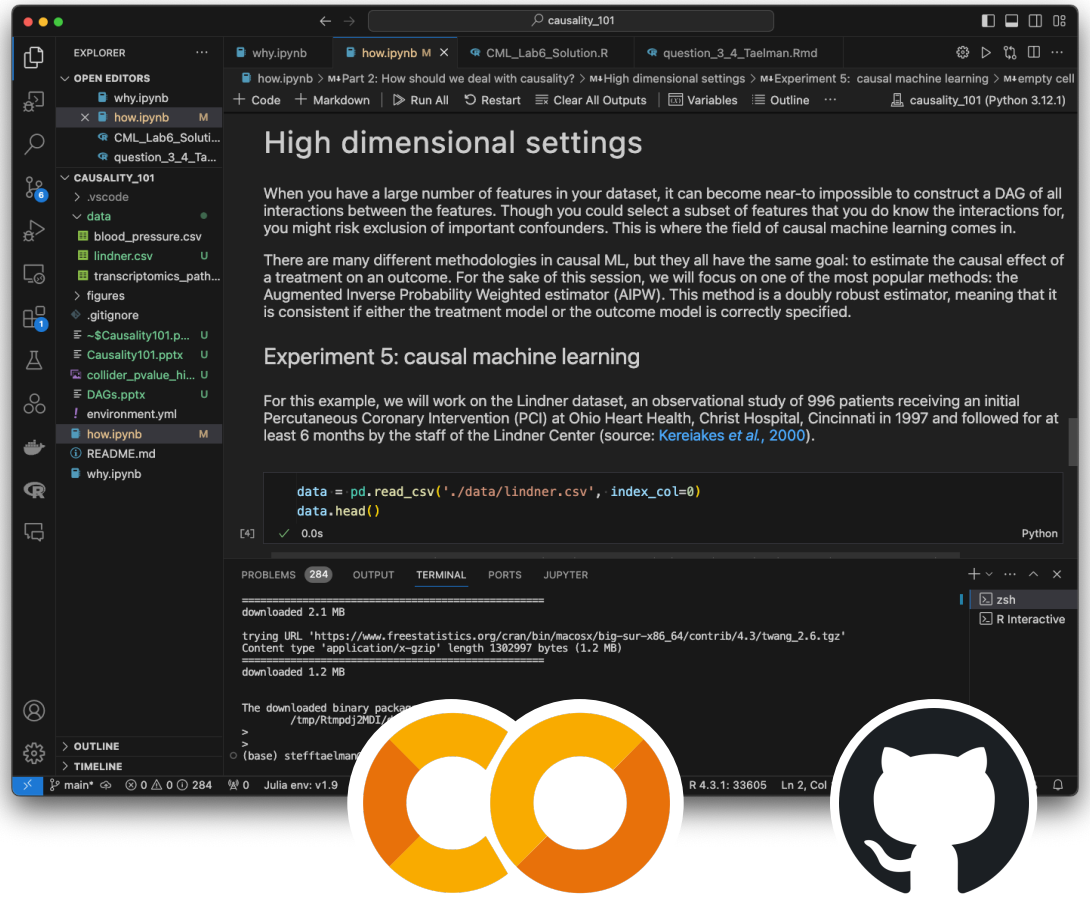
BioBix

KERMIT

BioLizard

Causal machine learning

how.ipynb



UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

BioBix

KERMIT

BioLizard

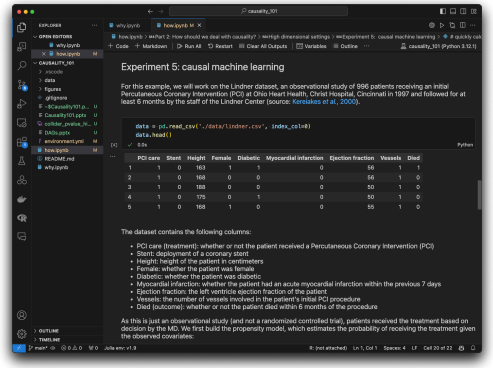
Causal machine learning - example

AHJ
American Heart Journal

Interventional Cardiology

Abciximab provides cost-effective survival advantage in high-volume interventional practice

Dean J. Kereiakes MD[a], Robert L. Obenchain PhD[b], Beth L. Barber PhD[b], Andrew Smith BS[a], Mark McDonald MA[c], Thomas M. Broderick MD[d], John Paul Runyon MD[d], Thomas M. Shimshak MD[d], John F. Schneider MD[d], Charles R. Hattemer MD[d], Eli M. Roth MD[d], David D. Whang MD[d], Douglas Cocks PhD[b], Charles W. Abbottsmith MD[d]



UNIVERSITEIT GENT

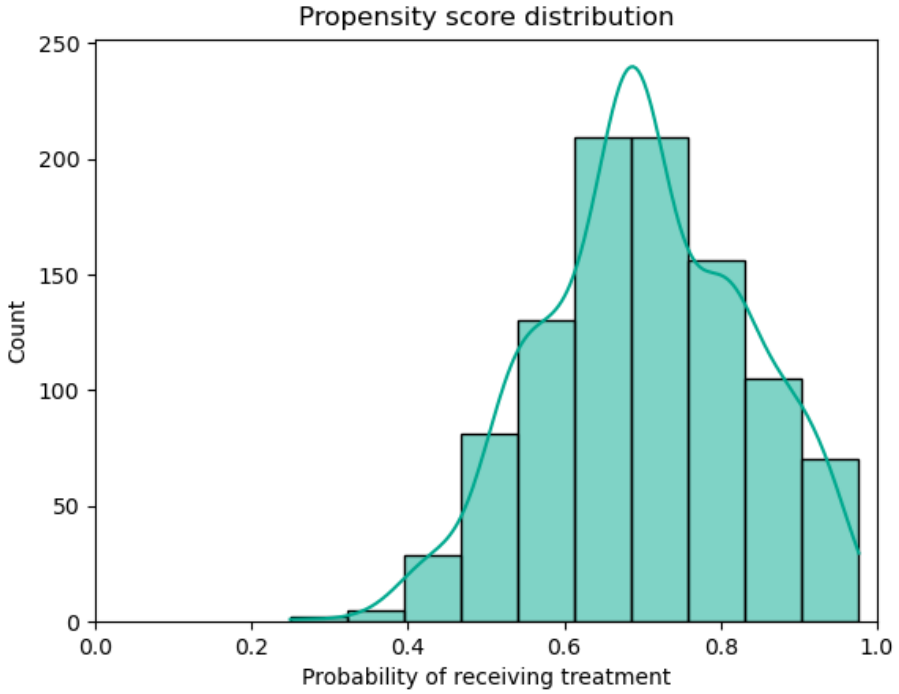
FACULTY OF BIOSCIENCE ENGINEERING

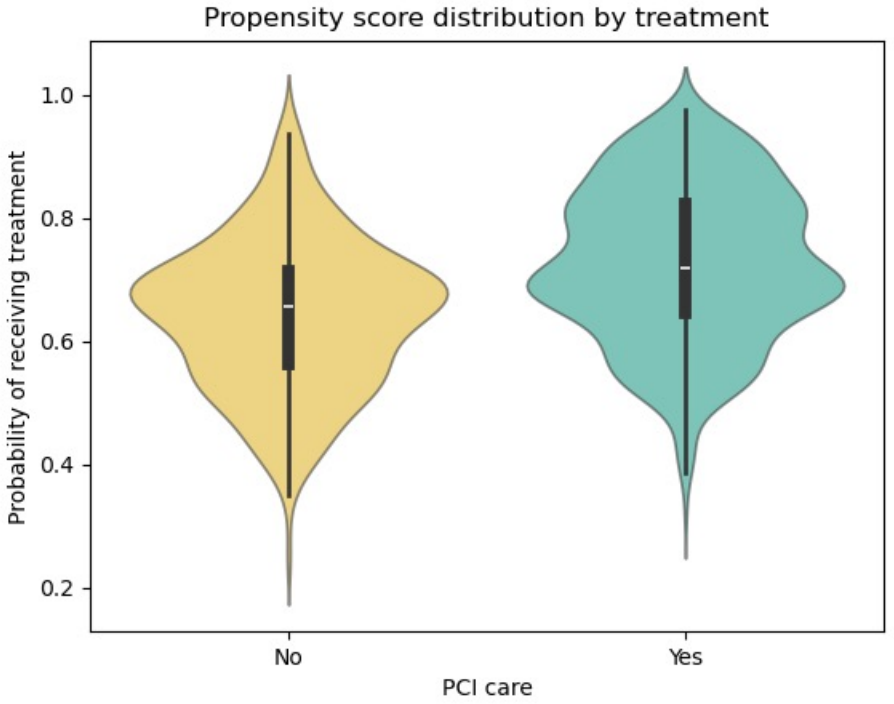
BioBix

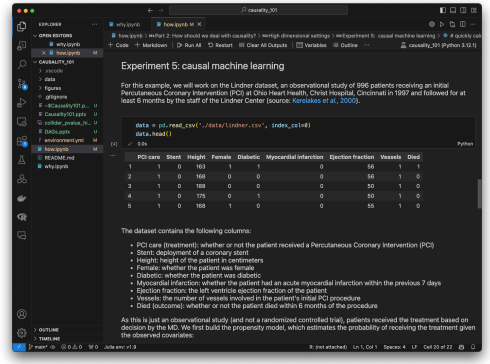
KERMIT

BioLizard

Causal machine learning - example







UNIVERSITEIT GENT

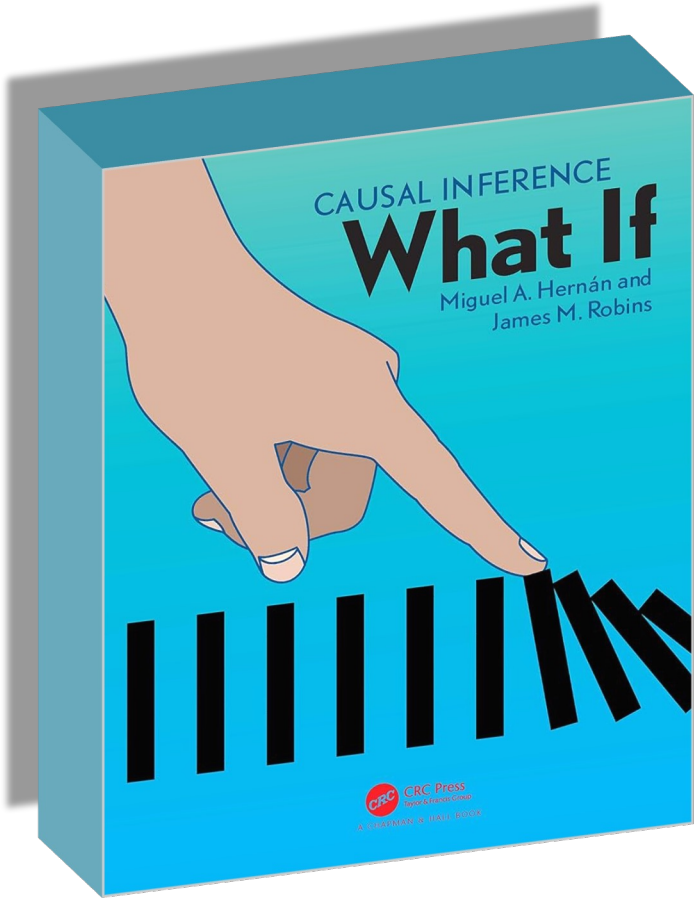
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

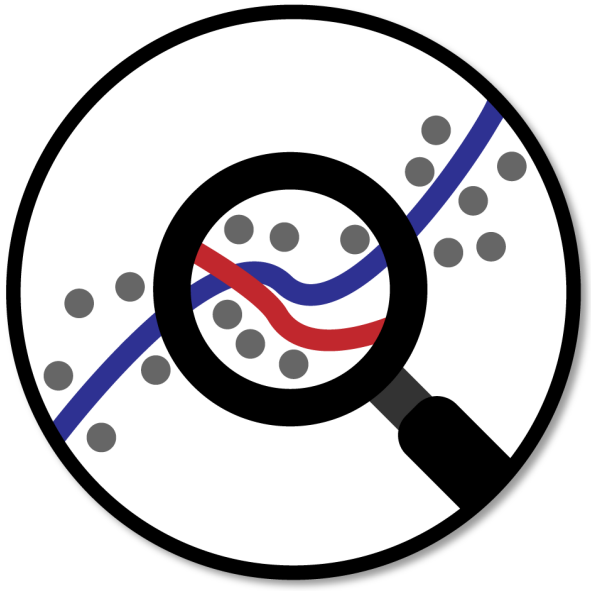
KERMIT

BioLizard

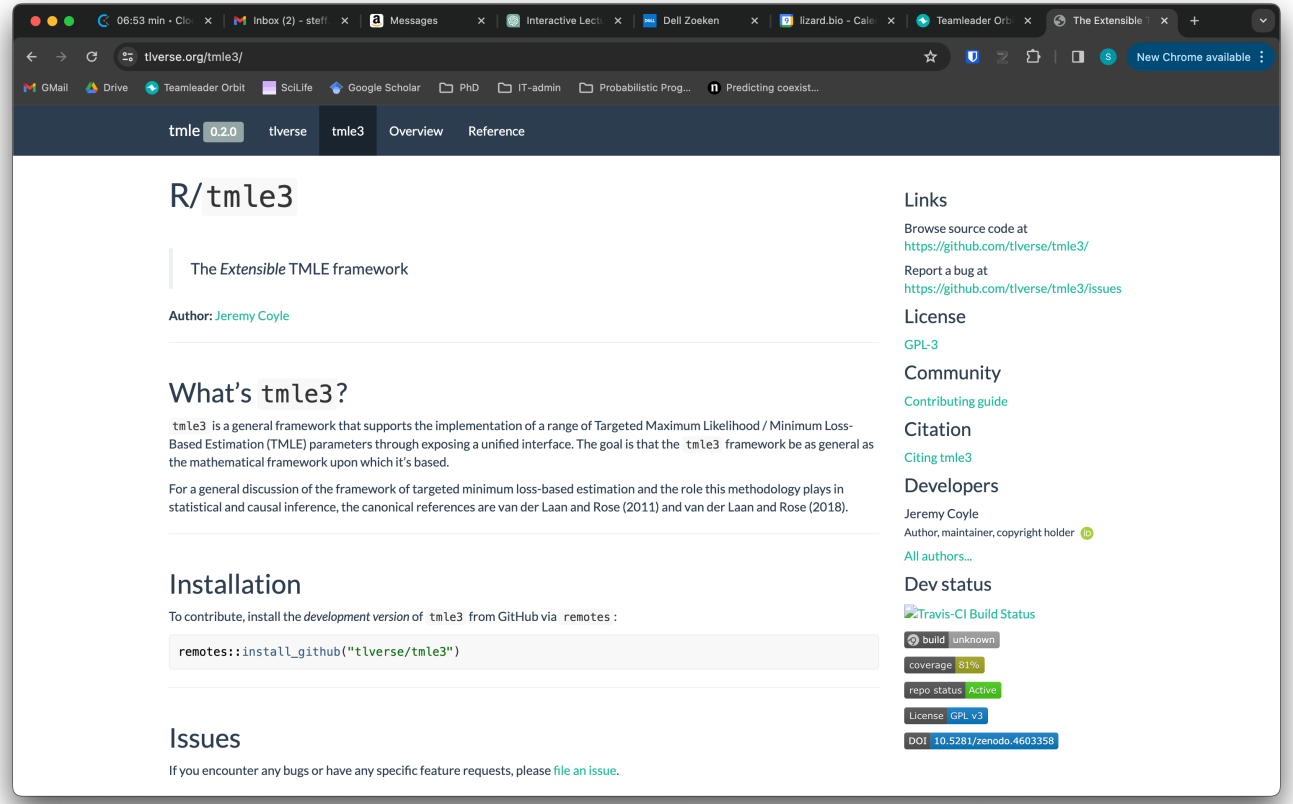
More about causal machine learning



More about causal machine learning



tlverse



UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

BioBix

KERMIT

BioLizard

Causality 101

How to draw causal statements from data and get away with it

Lizard Lab - 09.02.2024

UNIVERSITEIT GENT

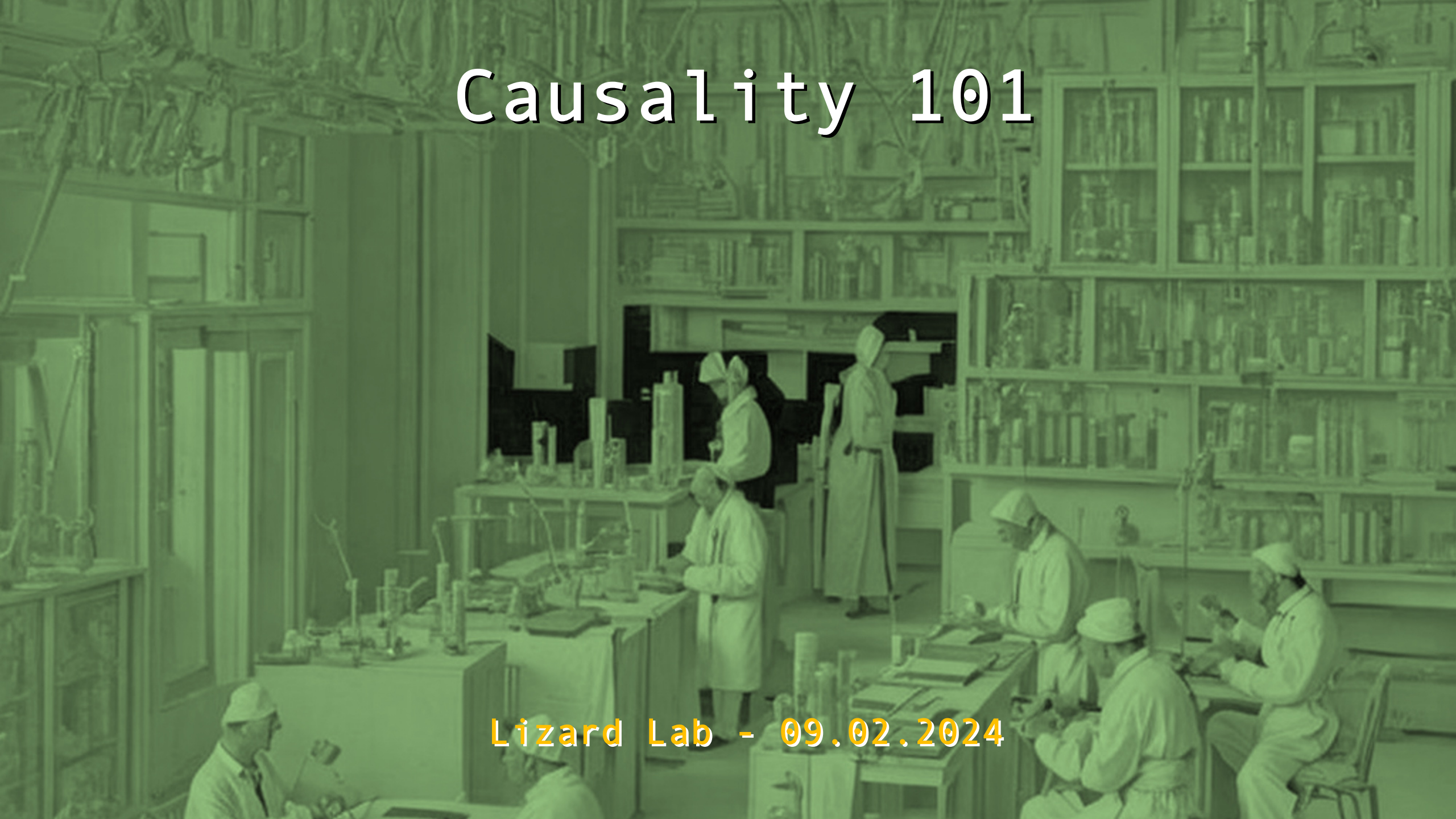
FACULTY OF BIOSCIENCE ENGINEERING

BioBix

KERMIT

BioLizard

Causality 101



Lizard Lab - 09.02.2024